

# PROPERTIES OF THE NONPARAMETRIC AUTOREGRESSIVE BOOTSTRAP

J. FRANKE, J.-P. KREISS, E. MAMMEN AND M. H. NEUMANN

**ABSTRACT.** We prove geometric ergodicity and absolute regularity of the nonparametric autoregressive bootstrap process. To this end, we revisit this problem for nonparametric autoregressive processes and give some quantitative conditions (i.e., with explicit constants) under which the mixing coefficients of such processes can be bounded by some exponentially decaying sequence. This is achieved by using well-established coupling techniques. Then we apply the result to the bootstrap process and propose some particular estimators of the autoregression function and of the density of the innovations for which the bootstrap process has the desired properties. Moreover, by using some “decoupling” argument, we show that the stationary density of the bootstrap process converges to that of the original process. As an illustration, we use the proposed bootstrap method to construct simultaneous confidence bands and supremum-type tests for the autoregression function as well as to approximate the distribution of the least squares estimator in a certain parametric model.

---

*Date:* October 26, 1998.

*1991 Mathematics Subject Classification.* Primary 62G09; secondary 62M10.

*Key words and phrases.* Bootstrap, nonparametric autoregression, coupling, geometric ergodicity, consistence.

## 1. INTRODUCTION

Since the seminal paper of Efron (1979), bootstrap methods have become a widely accepted and powerful tool to estimate the distribution as well as related quantities of certain statistics of interest. Typical fields of application are the construction of confidence sets for parameters or the closely related problem of determining the critical region for tests. The basic idea of the bootstrap in its original form is to mimic, on the basis of a single sample at hand, the whole structure of the data generating process. In the context of time series, this leads to the additional challenge of estimating the dependence structure of the process.

We assume throughout the present paper that data are generated by a nonparametric autoregressive process. Franke, Kreiss and Mammen (1997) discussed different bootstrap methods in this context. Besides two regression-type approaches including the wild bootstrap, they investigated the nonparametric autoregressive bootstrap which was first proposed by Franke and Wendel (1992) and Kreutzberger (1993), and proved its consistency for the pointwise behaviour of nonparametric estimators of the mean and the variance function. In subsequent papers, Neumann and Kreiss (1997) and Kreiss, Neumann and Yao (1998) showed the validity of the wild bootstrap beyond the pointwise distribution. The ultimate goal of the present paper is to open such a wide field of applications for the autoregressive bootstrap scheme. For this purpose, we first prove important basic properties of the bootstrap process such as absolute regularity and the convergence of the stationary distribution to that of the original process. Since the autoregressive bootstrap process is in particular a Markov chain, we can partially apply well-established techniques to prove the desired results. However, in contrast to many *qualitative* results in this field which simply state a certain rate for the decay of the mixing coefficients, we need here uniformity w.r.t. some parameters of the process varying within certain limits. This is because the properties of the bootstrap process depend on the original sample which is itself random. Hence, we will restate some well-known results with an explicit description of how constants depend on certain features of the process. To make the paper understandable for statisticians who are not specialists in Markov chain theory, we present self-contained versions of all major proofs.

These results can be used to prove consistency of the autoregressive bootstrap in several instances. We illustrate this by constructing simultaneous confidence bands and supremum-type tests for the autoregression function as well as by approximating the distribution of a least squares estimator in a certain parametric model.

## 2. MIXING OF MARKOV CHAINS REVISITED: A SET OF SUFFICIENT CONDITIONS FOR GEOMETRIC ERGODICITY

Throughout the present paper, our minimal assumption on the data generating process is that  $\{X_t\}$  forms a Markov chain. Properties like ergodicity and mixing are usually derived under two main assumptions: First, the existence of some “drift” towards a certain compact set  $K$ , and second, some condition on the conditional

distribution of future states, given that  $X_{t-1}$  falls into  $K$ . The latter condition ensures that information about previous states will be forgotten sufficiently fast by the Markov chain. Here is the first of our main conditions on the Markov chain:

**(A1)** There exists a compact set  $K$  such that

(i) there exist  $\rho > 1$  and  $\varepsilon > 0$  with

$$E(|X_t| | X_{t-1} = x) \leq \rho^{-1}|x| - \varepsilon \quad \text{for all } x \notin K,$$

(ii) there exists  $A < \infty$  with

$$\sup_{x \in K} \{E(|X_t| | X_{t-1} = x)\} \leq A.$$

The drift criterion already ensures that the set  $K$  is reached from every point with probability 1. However, it is not clear so far, which particular point in  $K$  is the first one visited by the Markov chain. If, for example,  $K$  contains more than one absorbing set, then it is *a priori* not clear to which of these sets the Markov chain will converge. Moreover, it might also happen that the Markov chain is periodic, that is, it moves periodically through a finite cycle of disjoint sets. There are well-known techniques to handle such cases, however, in order to facilitate the technical part of this paper, we will impose a condition that excludes them.

**(A2)** (i)  $K$  is a small set, that is,

there exist  $n_0 \in \mathbb{N}$ ,  $\gamma > 0$  and a probability measure  $\phi$  such that

$$\inf_{x \in K} \{P^{n_0}(x, B)\} \geq \gamma \phi(B)$$

holds for all measurable sets  $B$ .  $P^n(x, \cdot)$  denotes the  $n$ -step transition probability of the Markov chain started in  $x$ .

(ii) There exists  $\kappa > 0$  such that

$$\inf_{x \in K} \{P(x, K)\} \geq \kappa.$$

*Remark 1.*

- (i) Classical properties like irreducibility, aperiodicity and the existence of a unique stationary density follow readily from (A1) and (A2); see the proof of Theorem 2.1.
- (ii) To ensure aperiodicity and irreducibility, one often assumes instead of (A2) that the innovations,  $\varepsilon_t = X_t - m(X_{t-1})$ , are i.i.d. with an everywhere positive density. However, as noted by Meyn and Tweedie (1993, page 99), such a condition is unnecessarily restrictive. A possible condition which immediately implies (A2) and does not require an everywhere positive density of the innovations is the following one:

**(A2')** The conditional distribution  $\mathcal{L}(X_t | X_{t-1} = x)$  has a density  $p(y|x)$  which fulfills, for some  $c, \varepsilon > 0$ ,

$$p(y|x) \geq c > 0 \quad \text{for all } x, y \in K \text{ with } |x - y| \leq \varepsilon.$$

- (iii) Assumption (A2) allows the distribution of the innovations  $\varepsilon_t = X_t - m(X_{t-1})$  to depend on  $X_{t-1}$ , which in particular allows for conditional heteroscedasticity. We prove our results in this section in this general context, whereas we restrict them when dealing with the autoregressive bootstrap in the next section.
- (iv) If  $\{X_t\}$  can be written as  $X_t = m(X_{t-1}) + \varepsilon_t$ , where the innovations  $\varepsilon_t$  are i.i.d. with mean 0 and  $E|\varepsilon_t| < \infty$ , then (A1) follows from

$$\limsup_{|x| \rightarrow \infty} \{|m(x)/x|\} < 1.$$

The following lemma provides an important result about exponential moments of return times to  $K$ . The return time is defined as  $\tau_K = \inf\{t \geq 1 \mid X_t \in K\}$ . Moreover, we denote by  $E_x$  the conditional expectation under the condition that  $X_0 = x$ .

**Lemma 2.1.** *Suppose that (A1) is fulfilled. Then*

- (i)  $E_x \rho^{\tau_K} \leq \varepsilon^{-1}|x|$  for all  $x \notin K$ ,
- (ii)  $E_x \rho^{\tau_K} \leq \rho(1 + \varepsilon^{-1}A)$  for all  $x \in K$ .

Lemma 2.1 is the main tool to prove, in conjunction with assumption (A2), geometric ergodicity of the Markov chain, that is

$$\int \|P^n(x, \cdot) - \pi\|_{var} \mu(dx) \leq C_\mu \rho_\mu^{-n} \quad (2.1)$$

for some  $\rho_\mu > 1$ , where  $\|\cdot\|_{var}$  stands for the total variation norm and  $\mu$  stands for  $\pi$  if  $\{X_t\}$  is started with the stationary distribution  $\pi$ , or for the Dirac measure  $\delta_{x_0}$  if  $\{X_t\}$  is started at some nonrandom point  $x_0$ .

Exponential ergodicity will be proved via coupling of two Markov chains, one started at some nonrandom point  $x$ , and the other one started with initial distribution  $\pi$ . We pair both chains in such a way that they are completely identical to each other after they arrived at any state simultaneously. The coupling of  $\{X_t\}$  and  $\{X'_t\}$  is actually organized in two steps. Both chains are run independently until they reach the set  $K$  simultaneously, perhaps still at different points  $x$  and  $x'$ . By (A2), the set  $K$  is an appropriate place for an attempt to initiate an exact pairing which may occur after  $n_0$  further steps with a probability of at least  $\gamma$ . Lemma 2.1 guarantees, in conjunction with (A2)(ii), that a simultaneous entry in the set  $K$  occurs sufficiently often. This leads to the following theorem:

**Theorem 2.1.** *Suppose that (A1) and (A2) are fulfilled. Then (2.1) holds true with some  $\rho_\mu > 1$  and  $C_\mu < \infty$  which only depend on  $K, \rho, \varepsilon, A, n_0, \gamma, \kappa$ .*

Having proved geometric ergodicity, we obtain the desired geometric absolute regularity immediately from Proposition 1 of Davydov (1973). The coefficient of absolute regularity is defined as follows.

Let  $(\Omega, \mathcal{A}, P)$  be a probability space and let  $\mathcal{U}$  and  $\mathcal{V}$  be two  $\sigma$ -subfields of  $\mathcal{A}$ . The coefficient of absolute regularity ( $\beta$ -mixing coefficient) is defined as

$$\begin{aligned}\beta(\mathcal{U}, \mathcal{V}) &= E \left[ \sup_{V \in \mathcal{V}} \{|P(V | \mathcal{U}) - P(V)|\} \right] \\ &= \sup_{U_i \in \mathcal{U}, V_j \in \mathcal{V}} \frac{1}{2} \left\{ \sum_{i,j} |P(U_i)P(V_j) - P(U_i \cap V_j)| \right\},\end{aligned}$$

where the supremum in the last expression is taken over all finite partitions  $(U_i)_{i \in I}$  and  $(V_j)_{j \in J}$  of  $\Omega$  with  $U_i \in \mathcal{U}$  and  $V_j \in \mathcal{V}$ .

In our particular case of a possibly nonstationary process  $\{X_t\}_{t=0,1,\dots}$ , we adopt the definition of Davydov (1973), namely

$$\beta(s) = \sup_t E \left[ \sup_{B \in \mathcal{M}_{t+s}^\infty} |P(B | \mathcal{M}_0^t) - P(B)| \right],$$

where  $\mathcal{M}_u^v = \sigma(X_u, \dots, X_v)$ . [Note that Davydov had an additional factor of 2 in comparison with our definition of  $\beta(s)$ .]

The following lemma shows the close connection of ergodicity and absolute regularity for Markov chains.

**Lemma 2.2.** *(adapted from Davydov (1973))*

Let  $\{X_t\}$  be a Markov chain with marginal distributions  $X_t \sim \nu_t$ . Then

$$\beta(s) = \frac{1}{2} \sup_t \int \nu_t(dx) \|P^s(x, \cdot) - \nu_{t+s}\|_{Var}.$$

Now we obtain, in conjunction with Theorem 2.1, the desired mixing property of the Markov chain. Recall that  $\mu$  is used to denote the initial distribution, that is  $X_0 \sim \mu$ .

**Corollary 2.1.** *Suppose that (A1) and (A2) are fulfilled. Then*

$$\beta(n) \leq C_\mu \rho_\mu^{-n}.$$

So far we have derived sufficient conditions for geometric ergodicity in the general context of a Markov chain  $\{X_t\}$ . The nonparametric autoregressive bootstrap, which we study in the next section, is tailored for the special case that  $\{X_t\}$  can be written in the form of a nonparametric autoregressive model,

$$X_t = m(X_{t-1}) + \varepsilon_t, \tag{2.2}$$

where the innovations  $\varepsilon_t$  are independent, identically distributed random variables with mean 0. It can be easily seen that the following condition implies (A1) and (A2):

**(A3)**  $\{X_t\}$  obeys (2.2), where

- (i)  $|m(x)| \leq C_1 + C_2|x|$  for all  $x$  and some  $C_1 < \infty$ ,  $C_2 < 1$ ,
- (ii)  $E|\varepsilon_t| < \infty$ ,
- (iii)  $p_\varepsilon(x) \geq C_3 > 0$  for all  $x \in [-C_4 - \sup_{x \in K}\{m(x) - x\}, C_4 - \inf_{x \in K}\{m(x) - x\}]$  and some  $C_4 > 0$ , where  $K = [-C_5, C_5]$ ,  $C_5 > (C_1 + E|\varepsilon_t|)/(1 - C_2)$ .

### 3. THE NONPARAMETRIC AUTOREGRESSIVE BOOTSTRAP

In this section we will investigate important basic properties of the autoregressive bootstrap and therefore we restrict the quite general structure of the data generating process as considered in the previous section to the special case (2.2), where  $\varepsilon_t$  are i.i.d. with mean 0 and variance  $\sigma^2$ . To ensure mixing properties to hold for  $\{X_t\}$ , we assume that the  $\varepsilon_t$  have a density  $p_\varepsilon$ .

The nonparametric autoregressive bootstrap is a generalization of an idea of Efron and Tibshirani (1986) and Holbert and Son (1986) for the case of linear autoregression, and has been first proposed by Franke and Wendel (1992) and Kreutzberger (1993). It was proved in Franke *et al.* (1997) that this method is asymptotically consistent for the pointwise properties of kernel estimators of  $m$ . We continue this investigation and derive some important properties of this bootstrap method which will allow to apply this technique also for other problems such as the construction of simultaneous confidence bands and supremum-type tests for the autoregression function as well as for approximating the distribution of a least squares estimator in a certain parametric model.

**3.1. Some basic properties of the autoregressive bootstrap.** The implementation of the nonparametric autoregressive bootstrap requires explicit estimates  $\widehat{m}$  and  $\widehat{p}_\varepsilon$  of  $m$  and  $p_\varepsilon$ , respectively. Before we propose some particular estimators, we formulate quite general conditions that ensure ergodicity and absolute regularity of the bootstrap process as well as some consistency properties. The bootstrap process is generated according to the equation

$$X_t^* = \widehat{m}(X_{t-1}^*) + \varepsilon_t^*, \quad t = 1, \dots, T, \quad (3.1)$$

where the  $\varepsilon_t^*$  are i.i.d. with density  $\widehat{p}_\varepsilon$ . Under the conditions given below, there exists a stationary distribution  $\pi^*$ . For simplicity we assume that  $\{X_t^*\}$  is stationary, that is,  $X_0^* \sim \pi^*$ .

To prove ergodicity and absolute regularity of  $\{X_t^*\}$ , we need only some analog to (A3) for  $\widehat{m}$  and  $\widehat{p}_\varepsilon$  in place of  $m$  and  $p_\varepsilon$ , respectively. On the other hand, such a result alone would be of little use because one applies bootstrap methods to imitate some features of the original process. One of the minimal requirements is certainly that the stationary distribution of  $\{X_t^*\}$  approximates that of  $\{X_t\}$  in some appropriate sense. This will be ensured by suitable conditions on the consistency of the estimates  $\widehat{m}$  and  $\widehat{p}_\varepsilon$ . We make throughout this paper the convention that  $\delta > 0$  denotes an arbitrarily small and  $\lambda < \infty$  an arbitrarily large constant. Moreover, we use the letter  $\vartheta > 0$  to denote some appropriately chosen positive constant. Besides (A3),

we will assume

- (A4)** There exists an appropriate sequence of sets  $\Omega_T \subseteq \mathbb{R}^{T+1}$  with  $P((X_0, \dots, X_T) \notin \Omega_T) = o(1)$ , such that for  $(X_0, \dots, X_T) \in \Omega_T$  the following properties are fulfilled:
- (i)  $|\widehat{m}(x)/x| \leq C_1 + C_2|x|$ , for some  $C_1 < \infty$  and  $C_2 < 1$ . [W.l.o.g. we assume that  $C_1$  and  $C_2$  coincide with the constants in (A3).]
  - (ii)  $\sup_{x \in \mathcal{X}_T} \{|\widehat{m}(x) - m(x)|\} = O(T^{-\vartheta})$  for an appropriate sequence of sets  $\mathcal{X}_T \in \mathbb{R}$  with  $P(X_t \notin \mathcal{X}_T) = O(T^{-\vartheta})$ ,
  - (iii)  $\|\widehat{p}_\varepsilon - p_\varepsilon\|_\infty \leq CT^{-\vartheta}$ ,
  - (iv)  $\int |\widehat{p}_\varepsilon(x) - p_\varepsilon(x)| dx \leq CT^{-\vartheta}$ ,
  - (v) for all  $M$  there exists some  $C_M < \infty$  such that

$$\int |\varepsilon|^M \widehat{p}_\varepsilon(\varepsilon) d\varepsilon \leq C_M.$$

We propose in the next subsection particular estimators  $\widehat{m}$  and  $\widehat{p}_\varepsilon$  that satisfy (A4) under suitable conditions. Under (A3) and (A4),  $\widehat{m}$  and  $\widehat{p}_\varepsilon$  fulfil the conditions of (A3) (possibly with different constants) with high probability. Hence, according to Theorem 2.1,  $\{X_t^*\}$  is geometrically ergodic, which implies geometric absolute regularity. This is formalized in the following theorem:

**Theorem 3.1.** *Suppose that the data generating process obeys (2.2) and that (A3) and (A4) are fulfilled. Let  $\beta^*(n)$  be the coefficient of absolute regularity of the process  $\{X_t^*\}$ . Then there exists some  $\rho_b > 1$  such that*

$$\beta^*(n) \leq C_b \rho_b^{-n}$$

*holds if  $(X_0, \dots, X_T) \in \Omega_T$ .*

In the proofs of the previous theorems, we use coupling of Markov chains to get geometric ergodicity. To prove closeness of the stationary distributions of  $\{X_t\}$  and  $\{X_t^*\}$ , we use the opposite approach which we call decoupling: We start both chains at a common point,  $X_0 \equiv X_0^* \equiv x_0$ , and analyze the decoupling of appropriately paired versions of them. Since, according to (A4), the transition probabilities are similar, we can couple both chains in such a way that  $P(X_n \neq X_n^*)$  increases slowly. On the other hand, both chains are geometrically ergodic. Therefore,  $P^n(x_0, \cdot)$  and  $P^{*n}(x_0, \cdot)$  converge quite fast to  $\pi$  and  $\pi^*$ , respectively. This idea leads to the following theorem which characterizes the closeness of the respective stationary distributions  $\pi$  and  $\pi^*$ .

**Theorem 3.2.** *Suppose that the data generating process obeys (2.2) and that (A3) and (A4) are fulfilled. Then*

$$\sup_{B \text{ measurable}} \left\{ \left( \lambda(B) T^{-\vartheta} + T^{-\lambda} \right)^{-1} |\pi(B) - \pi^*(B)| \right\} \leq C$$

*holds if  $(X_0, \dots, X_T) \in \Omega_T$ , where  $\lambda(\cdot)$  denotes the Lebesgue measure.*

**3.2. Particular estimators of  $m$  and  $p_\varepsilon$ .** The consistency of the autoregressive bootstrap follows from suitable consistency properties of  $\widehat{m}$  and  $\widehat{p}_\varepsilon$ . Franke *et al.* (1997) proved an appropriate kind of uniform consistency of  $\widehat{m}$  on a sequence of sets  $[-\gamma_T, \gamma_T]$ ,  $\gamma_T \rightarrow \infty$ , under the additional assumption that the stationary density  $\pi$  is not less than  $c_T$  ( $c_T \rightarrow 0$  with a suitable rate) on  $[-\gamma_T, \gamma_T]$ . Here we try to avoid this condition and impose regularity conditions solely on  $m$  and  $p_\varepsilon$ . To be able to estimate  $m$  with a sufficient accuracy, we assume that

**(A5)**  $m$  is Lipschitz continuous.

To facilitate our proofs, in particular that of the consistency of a certain estimator of  $m$ , we assume that

**(A6)** All moments of  $\varepsilon_t$  are finite.

In contrast to regression-type methods such as the wild bootstrap, it is also important to estimate the distribution of the innovations  $\varepsilon_t$  consistently. We will assume that

**(A7)**  $p_\varepsilon$  is Lipschitz and of bounded total variation.

In view of the different size of the stationary density in different regions, it seems natural to use a nearest neighbor estimator of  $m$ , which is defined as

$$\widehat{m}_N(x) = N^{-1} \sum_{t: X_{t-1} \in \widehat{\mathcal{N}}_N(x)} X_t. \quad (3.2)$$

The (random) neighborhoods  $\widehat{\mathcal{N}}_N(x) = [x - \widehat{n}_N(x), x + \widehat{n}_N(x)]$  are chosen such that  $\#\{t \leq T \mid X_{t-1} \in \widehat{\mathcal{N}}_N(x)\} = N$ , where  $N = N(T) \rightarrow \infty$  as  $T \rightarrow \infty$ . Instead of  $\widehat{m}_N$  one could also use other nonparametric estimators such as kernel or local polynomial estimators with appropriate adjustments of the bandwidths in regions of a low stationary density.

Since many assertions in this article are of the type that a certain random variable is below some threshold with a high probability, we introduce the following notation.

**Definition 3.1.** Let  $\{Z_T\}$  be a sequence of random variables and let  $\{\alpha_T\}$  and  $\{\gamma_T\}$  be sequences of positive reals. We write

$$Z_T = \widetilde{O}(\alpha_T, \gamma_T),$$

if

$$P(|Z_T| > C\alpha_T) \leq C\gamma_T$$

holds for  $T \geq 1$  and some  $C < \infty$ .



This definition is obviously stronger than the usual  $O_P$  and it is well suited for our particular purposes of constructing confidence bands and nonparametric tests; see its application in Section 4.

The following lemma provides a useful result about the uniform convergence properties of  $\widehat{m}_N$ .

**Lemma 3.1.** *Suppose that the data generating process obeys (2.2) and that (A3), (A5) and (A6) are fulfilled. Then there exists a sequence of sets  $\mathcal{X}_T \in \mathbb{R}$  with  $P(X_t \notin \mathcal{X}_T) = O(T^{-\vartheta})$  and*

$$\sup_{x \in \mathcal{X}_T} \{|\widehat{m}_N(x) - m(x)|\} = \tilde{O}\left(T^\delta N/T + N^{-1/2} \log(T), T^{-\lambda}\right).$$

Define

$$\widehat{p}_\varepsilon(x) = \frac{1}{T} \sum_{t=1}^T \frac{1}{h} K\left(\frac{x - \widehat{\varepsilon}_t}{h}\right),$$

where  $\widehat{\varepsilon}_t = X_t - \widehat{m}_N(X_{t-1})$  are the residuals.

**Lemma 3.2.** *Suppose that the data generating process obeys (2.2) and that (A3) and (A5) to (A7) are fulfilled. Furthermore, let  $h$  and  $N$  be chosen such that  $h = O(T^{-\vartheta'})$ ,  $h^{-1} = O(N^{1/2}T^{-\vartheta'})$  and  $N = O(T^{1-\vartheta'})$  for some  $\vartheta' > 0$ . Then there exists some  $\vartheta > 0$  such that*

- (i)  $\|\widehat{p}_\varepsilon - p_\varepsilon\|_\infty = \tilde{O}\left(T^{-\vartheta}, T^{-\lambda}\right),$
- (ii)  $\int |\widehat{p}_\varepsilon(x) - p_\varepsilon(x)| dx = \tilde{O}\left(T^{-\vartheta}, T^{-\lambda}\right).$

#### 4. APPLICATION TO PARAMETRIC AND NONPARAMETRIC ESTIMATES OF THE AUTOREGRESSION FUNCTION

In the first part of this section we use the proposed bootstrap method to construct simultaneous confidence bands and supremum-type tests for the autoregression function. Similar results for a regression-type bootstrap, the so-called wild bootstrap, can be found in Neumann and Kreiss (1997). The validity of the wild bootstrap in context with nonparametric estimation in autoregression relies on the fact that the underlying statistic forms a sum of martingale differences. Moreover, bootstrap methods based on the (fictive) assumption of independent random variables are consistent for many statistics based on nonparametric estimators in the context of general processes since the effect of weak dependence vanishes asymptotically; see, e.g., Neumann (1996, 1997). Usually, this is not true for parametric estimation. In such a situation a process bootstrap as proposed in this paper is really necessary for consistency, since the whole dependence structure of the underlying process has to be mimicked. One may argue that this may motivate the use of process bootstrap even for nonparametric estimation. However, for nonparametric estimation, a rigorous

comparison of process bootstrap with other resampling schemes would require higher order methods.

**4.1. Application to supremum-type statistics: confidence bands and goodness-of-fit tests.** We suppose that the data generating process obeys (2.1). A simultaneous confidence band for  $m$  is usually based on and centered around some nonparametric estimator  $\widehat{m}_h(x)$ . For simplicity, one can take a Nadaraya-Watson kernel estimator,

$$\widehat{m}_h(x) = \frac{\sum_{t=1}^T K\left(\frac{x-X_{t-1}}{h}\right) X_{t-1}}{\sum_{t=1}^T K\left(\frac{x-X_{t-1}}{h}\right)}. \quad (4.1)$$

The difference of  $\widehat{m}_h(x)$  and  $m(x)$  can be decomposed into a stochastic term,

$$\left(\sum_t K((x - X_{t-1})/h)\right)^{-1} \sum_t K((x - X_{t-1})/h) [X_t - m(X_{t-1})], \quad (4.2)$$

and a bias-type term,

$$\left(\sum_t K((x - X_{t-1})/h)\right)^{-1} \sum_t K((x - X_{t-1})/h) m(X_{t-1}) - m(x). \quad (4.3)$$

[We call the latter expression “bias-type term” rather than “bias term” since it is only asymptotically nonrandom.]

For the construction of confidence intervals or bands, one may account for the bias-type term by separate adjustments, i.e., it is not necessary to imitate it by the bootstrap. Usual techniques are undersmoothing or explicit bias correction; see, e.g., Neumann and Kreiss (1997) for a discussion in the context of nonparametric autoregression. In order to find an appropriate width of the confidence band, it remains to get knowledge about the stochastic term. This term can be approximated by  $[(p * K(. / h))(x)]^{-1} 1/T \sum_t K((x - X_{t-1})/h) \varepsilon_t$ , where  $p$  denotes the density of  $\pi$ . Hence, we have to approximate the distribution of

$$S_T = \sup_{x \in [a, b]} \left\{ [(p * K(. / h))(x)]^{-1} \left| \frac{1}{T} \sum_t K\left(\frac{x - X_{t-1}}{h}\right) \varepsilon_t \right| \right\}. \quad (4.4)$$

For a parametric hypothesis  $H_0 : m \in \mathcal{M} = \{m_\theta \mid \theta \in \Theta\}$  we can use the test statistic

$$W_T = \sup_{x \in \mathbb{R}} \left\{ \left| \sum_t K\left(\frac{x - X_{t-1}}{h}\right) [X_t - \widehat{m}(X_{t-1})] \right| \right\}, \quad (4.5)$$

where  $\widehat{m}$  is any estimator that satisfies on the hypothesis  $m \in \mathcal{M}$

$$\sup_{x \in \mathbb{R}} \left\{ \left| \sum_t K\left(\frac{x - X_{t-1}}{h}\right) [\widehat{m}(X_{t-1}) - m(X_{t-1})] \right| \right\} = o_P((Th)^{1/2}(\log T)^{-1/2}). \quad (4.6)$$

For the determination of critical values we have to approximate the distribution of  $W_T$ . A sufficient condition for (4.6) to be fulfilled is obviously that  $\widehat{m}$  itself

converges on the hypothesis in the supremum norm to  $m$  with a faster rate than  $(Th)^{-1/2}(\log T)^{-1/2}$ . If (4.6) is actually satisfied, it suffices to find a consistent approximation to the distribution of the statistic

$$U_T = \sup_{x \in \mathbb{R}} \left\{ \left| \sum_t K \left( \frac{x - X_{t-1}}{h} \right) \varepsilon_t \right| \right\}, \quad (4.7)$$

which is closely related to  $S_T$  in (4.4).

The distributions of  $S_T$  and  $U_T$  can be approximated by those of appropriate bootstrap statistics. We discuss only the approximation of  $U_T$  by

$$U_T^* = \sup_{x \in \mathbb{R}} \left\{ \left| \sum_t K \left( \frac{x - X_{t-1}}{h} \right) \varepsilon_t^* \right| \right\} \quad (4.8)$$

more closely. Whereas a purely analytic approach of showing consistency is presumably quite cumbersome for such supremum-type functionals, a proof via strong approximations is much more convenient.

**Lemma 4.1.** *Suppose that the data generating process obeys (2.2) and that (A3) and (A4) are fulfilled. Then there exists on a sufficiently large probability space a pairing of  $(X_0, \varepsilon_1, \dots, \varepsilon_T)$  and  $(X_0^*, \varepsilon_1^*, \dots, \varepsilon_T^*)$  such that*

$$\sup_{x \in \mathbb{R}} \left\{ \left| \sum_t K \left( \frac{x - X_{t-1}}{h} \right) \varepsilon_t - \sum_t K \left( \frac{x - X_{t-1}^*}{h} \right) \varepsilon_t^* \right| \right\} = o_P \left( (Th)^{1/2} (\log T)^{-1/2} \right)$$

*holds uniformly over all bootstrap distributions  $\mathcal{L}((X_0^*, \varepsilon_1^*, \dots, \varepsilon_T^*) \mid X_0, \dots, X_T)$  for  $(X_0, \dots, X_T) \in \Omega_T$ , where  $\Omega_T$  is an appropriate set with  $P(\Omega_T^c) = o(1)$ .*

This strong approximation result basically says that the stochastic behaviour of the process  $\{\sum_t K((x - X_{t-1})/h) \varepsilon_t\}_{x \in \mathbb{R}}$  is well approximated by that of the bootstrap counterpart  $\{\sum_t K((x - X_{t-1}^*)/h) \varepsilon_t^*\}_{x \in \mathbb{R}}$ . This implies in particular that the distribution of  $U_T$  is consistently approximated by that of  $U_T^*$ . As can be seen from Lemma 3.2 in Neumann and Kreiss (1997), the rate of  $o_P((Th)^{1/2}(\log T)^{-1/2})$  for the approximation error is just sufficient for the validity of the bootstrap in the context of supremum-type functionals. Hence, we may apply the nonparametric autoregressive bootstrap to determine the critical value for a supremum-type test based on  $W_T$ . For the same reason it can also be used for the construction of simultaneous confidence bands.

**4.2. Application to a problem of parametric inference.** As an illustration for a situation where the nonparametric autoregressive bootstrap procedure (cf. Section 3) is really necessary, we consider the following example. Suppose that we intend to fit a parametric model,

$$X_t = m_\theta(X_{t-1}) + \varepsilon_t,$$

to the time series. For the sake of simplicity, let us deal with the simplest case, i.e.,  $m_\theta(u) = \theta \cdot m_o(u)$  for some known function  $m_o$  and the least squares estimator  $\hat{\theta}$ ,

which satisfies

$$\sqrt{T}(\hat{\theta} - \theta) = \frac{\frac{1}{\sqrt{T}} \sum_t (X_t - \theta m_o(X_{t-1})) m_o(X_{t-1})}{\frac{1}{T} \sum_t m_o(X_{t-1})^2}.$$

Recall that we do not assume that the parametric model coincides with the underlying model. If we assume (A1), (A2), (A3)(i) for  $m_o$  and  $E|X_t|^\gamma < \infty$  for some  $\gamma > 4$ , then we obtain from a CLT for strongly mixing processes, cf. Bosq (1996, Theorem 1.7), asymptotic normality for the least squares estimator  $\hat{\theta}$ , namely

$$\sqrt{T}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}\left(0, \tau^2 / (Em_o(X_0)^2)^2\right).$$

In the case of model inadequacy, the parameter  $\theta$  is defined in the sense of the best approximation, that is

$$\theta = \arg \min_{\tilde{\theta}} \left\{ E \left( X_1 - \tilde{\theta} m_o(X_0) \right)^2 \right\} = \frac{EX_1 m_o(X_0)}{Em_o(X_0)^2}.$$

The term

$$\begin{aligned} \tau^2 &= E(X_1 - \theta m_o(X_0))^2 m_o(X_0)^2 \\ &\quad + 2 \cdot \sum_{k=1}^{\infty} \text{Cov}[(X_1 - \theta m_o(X_0)) m_o(X_0), (X_{k+1} - \theta m_o(X_k)) m_o(X_k)] \\ &= E\varepsilon_1^2 \cdot Em_o(X_0)^2 + E(m(X_0) - \theta m_o(X_0))^2 m_o(X_0)^2 \\ &\quad + 2 \cdot \sum_{k=1}^{\infty} \text{Cov}[(X_1 - \theta m_o(X_0)) m_o(X_0), (X_{k+1} - \theta m_o(X_k)) m_o(X_k)] \end{aligned}$$

depends on the whole dependence structure of the process. The application of the wild bootstrap will lead in any case to an asymptotic normal distribution with variance  $E\varepsilon_1^2 / Em_o(X_0)^2$  which is in general not equal to  $\tau^2 / (Em_o(X_0)^2)^2$ . In contrast, the process bootstrap described in Section 3 leads to consistency. This is the content of the following result.

**Lemma 4.2.** *Suppose that the data generating process obeys (2.2) and that (A3), (A4) and (A7) are fulfilled. Then*

$$\sqrt{T} \left( \frac{\sum_t X_t^* m_o(X_{t-1}^*)}{\sum_t m_o(X_{t-1}^*)^2} - \theta^* \right) \Rightarrow \mathcal{N}\left(0, \tau^2 / (Em_o(X_0)^2)^2\right),$$

*holds if  $(X_0, \dots, X_T) \in \Omega_T$ .  $\theta^*$  denotes the value of the optimal fit (in the  $L_2$ -sense) of a parametric model in the bootstrap world, i.e.,  $\theta^* = E^* X_1^* m_o(X_0^*) / E^* m_o(X_0^*)^2 \rightarrow EX_1 m_o(X_0) / Em_o(X_0)^2 = \theta$  as  $T \rightarrow \infty$ .*

## 5. PROOFS

*Proof of Lemma 2.1.* A condensed proof of this lemma has already been given in Nummelin and Tuominen (1982).

(i) Let  $x \notin K$ . We get immediately from (A1)(i)

$$|x| - \rho E(|X_1| \mid X_0 = x) \geq \rho \varepsilon. \quad (5.1)$$

Analogously we have

$$I(y \in K^c) [|y| - \rho E(|X_2| \mid X_1 = y)] \geq \rho \varepsilon I(y \in K^c).$$

Multiplying both sides with  $\rho$  and taking the expectation over  $X_1$  under the condition  $X_0 = x$ , we obtain

$$E_x I(X_1 \in K^c) [\rho |X_1| - \rho^2 |X_2|] \geq \rho^2 \varepsilon P_x(X_1 \in K^c). \quad (5.2)$$

By analogous considerations, we get

$$E_x I(X_1, \dots, X_k \in K^c) [\rho^k |X_k| - \rho^{k+1} |X_{k+1}|] \geq \rho^{k+1} \varepsilon P_x(X_1, \dots, X_k \in K^c). \quad (5.3)$$

Now we obtain from (5.1) to (5.3) that

$$|x| \geq \varepsilon \sum_{k=0}^{\infty} \rho^{k+1} P_x(X_1, \dots, X_k \in K^c) = \varepsilon \sum_{k=0}^{\infty} \rho^{k+1} P_x(\tau_K > k) \geq \varepsilon E_x \rho^{\tau_K}.$$

(ii) For  $x \in K$ , we obtain that

$$\begin{aligned} E_x \rho^{\tau_K} &= \rho \int_K P(x, dy) + \rho \int_{K^c} P(x, dy) E_y \rho^{\tau_K} \\ &\leq \rho \left[ P_x(\tau_K = 1) + \varepsilon^{-1} \int_{K^c} P(x, dy) |y| \right]. \end{aligned}$$

[Notice that the term “ $P_x(\tau_K = 1)$ ” was missing in Theorem 3.1 of Nummelin and Tuominen (1982) as well as on page 90 in Doukhan (1994).]  $\square$

*Proof of Theorem 2.1.* (i) *Some preliminaries: Irreducibility, recurrence and the existence of  $\pi$*

First we check irreducibility of  $\{X_t\}$  since this simplifies the analysis by excluding the case of more than one absorbing set. By Lemma 2.1,  $\varphi = \lambda(\cdot \cap K)$  is obviously an irreducibility measure. According to Proposition 4.2.2 from Meyn and Tweedie (1993, p. 88), there also exists a maximal irreducibility measure  $\psi$ .

Since  $K$  is a small set with  $P_x(\tau_K < \infty) = 1$  for all  $x$ , we obtain from Theorem 8.3.6 in Meyn and Tweedie (1993, p. 187) that  $\{X_t\}$  is recurrent. ( $\{X_t\}$  is called *recurrent* if it is  $\psi$ -irreducible and  $\sum_{n=1}^{\infty} P^n(x, A) = \infty$  for each  $x \in \mathbb{R}$  and every measurable set  $A$  with  $\psi(A) > 0$ .)

Since  $\{X_t\}$  is recurrent, we conclude from Theorem 10.4.4 of Meyn and Tweedie (1993, p. 242) that there exists a unique invariant measure which we denote by  $\pi$ .

(ii) *Coupling*

Our proof of geometric ergodicity is mainly based on an appropriate coupling of one Markov chain started in some state  $x$  with another chain having an initial distribution equal to  $\pi$ . This is one of the classical approaches to prove ergodicity of Markov chains; see, for example, Lindvall (1992) and Meyn and Tweedie (1993). The most substantial novelty of our proof is that we focus on explicit constants which are necessary in view of the randomness of the parameters of the bootstrap process. Coupling consists of establishing an appropriate pairing of two Markov chains,

$$X_0, X_1, \dots \quad \text{with } X_0 \equiv x$$

and

$$X'_0, X'_1, \dots \quad \text{with } X'_0 \sim \pi,$$

on a joint probability space. Let  $\tau$  be the first time that both chains reach any state simultaneously. By the Markov property, we can pair these chains in such a way that  $X_t \equiv X'_t$  for all  $t \geq \tau$ . We call the time  $\tau$  the coupling time of the two processes. It is easy to see that

$$\begin{aligned} \|P^n(x, \cdot) - \pi\|_{Var} &= \sup_{f: |f| \leq 1} \left\{ \left| \int P^n(x, dy) f(y) - \int \pi(dy) f(y) \right| \right\} \\ &\leq 2P_x(X_n \neq X'_n) = 2P_x(\tau > n). \end{aligned} \quad (5.4)$$

For Markov chains with an accessible atom  $\alpha$  (A set  $\alpha$  is called an *atom* if there exists a probability measure  $\nu$  such that  $P(x, B) = \nu(B)$  for all  $x \in \alpha$ .) the construction of such a pairing is not difficult: One simply lets run both chains independently until they reach  $\alpha$  simultaneously, and from that time both chains are identical.

In our context, which includes the case of purely continuous distributions of the innovations  $\varepsilon_t$ , the existence of an accessible atom is not guaranteed. However, under assumption (A2)(i), we may use the splitting device of Nummelin (1978) and Athreya and Ney (1978) to introduce an appropriate substitute, which we also denote by  $\alpha$ , and which is an atom for the  $n_0$ -skeleton for the chain, that is

$$P^{n_0}(x, B) = \nu(B) \quad \text{for all } x \in \alpha, B \text{ measurable.}$$

Hence, we can couple  $\{X_t\}$  and  $\{X'_t\}$  in such a way that  $X_t \equiv X'_t$  for all  $t \geq \tau_{\alpha\alpha} + n_0$ , where  $\tau_{\alpha\alpha}$  is the time of the first common visit to the state  $\alpha$ .

To define a suitable substitute for the atom  $\alpha$ , we apply the idea of Athreya and Ney (1978) and use an additional randomization with the aid of independent random variables  $N_t$  and  $N'_t$ ,  $t = 1, \dots, T$ , with  $P(N_t = 1) = P(N'_t = 1) = \gamma$ . If  $X_t$  hits  $K$ , then we define the  $n_0$ -step transition probability equal to  $\nu(\cdot)$  if  $N_t = 1$  and equal to  $[P^{n_0}(X_t, \cdot) - \gamma\nu(\cdot)]/(1 - \gamma)$  if  $N_t = 0$ . (The same is done for the chain  $\{X'_t\}$  in dependence on the value of  $N'_t$ .) In other words,  $X_t$  hits the atom  $\alpha$  if  $X_t \in K$  and  $N_t = 1$ .

(iii) *An experiment consisting of successive trials*

In view of (5.4), it remains to find a pairing of  $\{X_t\}$  and  $\{X'_t\}$  such that

$$\int P_x(\tau_{\alpha\alpha} + n_0 > n) \mu(dx) \leq C_\mu \rho_\mu^{-n}, \quad (5.5)$$

where  $P_x$  refers to an initial condition of  $X_0 = x$  for the Markov chain.

To bound the probability  $P_x(\tau_{\alpha\alpha} + n_0 > n)$ , we consider successive trials of the chains  $\{X_t\}$  and  $\{X'_t\}$  to hit the state  $\alpha$  at the same time. We define stopping times  $\tau_i$  and  $\tau'_i$  that refer to certain events that  $\{X_t\}$  and  $\{X'_t\}$  visit  $K$ . Let

$$\tau_0 = \min_j \{X_j \in K\}$$

and

$$\tau'_0 = \min_j \{X'_j \in K \mid j \geq \tau_0\}.$$

Further, we define inductively

$$\tau_i = \min_j \{X_j \in K \mid j \geq \tau'_{i-1}\}$$

and

$$\tau'_i = \min_j \{X'_j \in K \mid j \geq \tau_i\}.$$

It is clear that  $\tau_i$  and  $\tau'_i$  are indeed stopping times with respect to the  $\sigma$ -field  $\mathcal{B}_i = \sigma(X_0, \dots, X_{\tau_i}, X'_0, \dots, X'_{\tau'_i})$ . These stopping times are defined in such a way that

$$0 \leq \tau_0 \leq \tau'_0 \leq \tau_1 \leq \tau'_1 \leq \dots \leq \tau_i \leq \tau'_i \leq \dots$$

The time  $\tau_{\alpha\alpha}$  corresponds to the first joint visit of the Markov chains  $\{X_t\}$  and  $\{X'_t\}$  at  $\alpha$ . Accordingly, we call a trial  $(\tau_i, \tau'_i, N_i, N'_i)$  successful if  $\tau_i = \tau'_{i-1}$ ,  $N_i = N'_{i-1} = 1$  or  $\tau'_i = \tau_i$ ,  $N'_i = N_i = 1$ . Our next step consists of showing that the conditional probability of a success of a trial  $(\tau_i, \tau'_i, N_i, N'_i)$  given  $\mathcal{B}_{i-1}$  is bounded away from zero. [Actually, we were not able to prove that  $P(\tau_i = \tau'_{i-1}, N_i = N'_{i-1} = 1 \mid \mathcal{B}_{i-1})$  can be bounded in such a way. It might happen that  $\tau'_{i-1} - \tau_{i-1}$  is arbitrarily large. Since we do not have an explicit lower bound for  $\inf_{j \geq L} \inf_{x \in \alpha} P_x^j(K)$ , we cannot derive an explicit lower bound for  $P(\tau_i = \tau'_{i-1}, N_i = N'_{i-1} = 1 \mid \mathcal{B}_{i-1})$ .] However, fortunately, we can find such a bound for  $P(\tau'_i = \tau_i, N'_i = N_i = 1 \mid \mathcal{B}_{i-1})$ . This explains why we are considering such “double trials”  $(\tau_i, \tau'_i, N_i, N'_i)$  rather than single trials.

Using the last-exit representation, we find by (ii) of Lemma 2.1 that

$$\begin{aligned} & P(\tau_i - \tau'_{i-1} \geq L \mid \mathcal{B}_{i-1}) \\ &= \sum_{s=0}^{\tau'_{i-1} - \tau_{i-1}} \int_K P_{X_{\tau_{i-1}}}^{\tau'_{i-1} - \tau_{i-1} - s}(dy) P_y(\tau_K \geq L + s) \\ &\leq \sum_{t=L}^{\infty} \sup_{y \in K} \{P_y(\tau_K \geq t)\} \\ &\leq \sum_{t=L}^{\infty} \sup_{y \in K} \{E_y \rho^{\tau_K - t}\} \leq C \sum_{t=L}^{\infty} \rho^{-t} \longrightarrow 0 \quad \text{as } L \rightarrow \infty. \end{aligned} \quad (5.6)$$

Hence,

$$P\left(\tau_i - \tau'_{i-1} \leq L_0 \mid \mathcal{B}_{i-1}\right) \geq C$$

for appropriate  $C > 0$  and  $L_0 < \infty$ , which implies that

$$\begin{aligned} & P(\text{the } i\text{th trial is successful} \mid \mathcal{B}_{i-1}) \\ & \geq P(\tau'_i = \tau_i, N'_i = N_i \mid \mathcal{B}_{i-1}) \\ & \geq \gamma^2 \sum_{j=0}^{L_0} P\left(\tau_i - \tau'_{i-1} = j, X'_{\tau'_{i-1}+j} \in K \mid \mathcal{B}_{i-1}\right) \\ & = \gamma^2 \sum_{j=0}^{L_0} P\left(\tau_i - \tau'_{i-1} = j \mid \mathcal{B}_{i-1}\right) P\left(X'_{\tau'_{i-1}+j} \in K \mid X'_{\tau'_{i-1}}\right) \\ & \geq \gamma^2 P\left(\tau_i - \tau'_{i-1} \leq L_0 \mid \mathcal{B}_{i-1}\right) \inf_{0 \leq j \leq L_0} \inf_{x \in K} \left\{P^j(x, K)\right\} \\ & \geq \tilde{\gamma} > 0. \end{aligned} \tag{5.7}$$

This is just the point where (A2)(ii), which is slightly stronger than aperiodicity of the Markov chain, enters into the proof of geometric ergodicity.

(iv) *An exponential bound for the coupling time*

Let  $\eta$  denote the number of the first successful trial  $(\tau_i, \tau'_i, N_i, N'_i)$ , that is,  $\tau_{\alpha\alpha} = \tau_\eta$ . We split up:

$$P(\tau_{\alpha\alpha} + n_0 > n) \leq P(\eta > [\kappa n]) + P(\tau_{[\kappa n]} + n_0 > n), \tag{5.8}$$

where the constant  $\kappa$  will be specified below.

Let  $T_i$  be the indicator of the event that the  $i$ th trial is successful, that is,

$$T_i = \begin{cases} 1 & \text{if } (\tau_i = \tau'_{i-1}, N_i = N'_{i-1}) \text{ or } (\tau'_i = \tau_i, N'_i = N_i) \\ 0 & \text{otherwise} \end{cases}.$$

According to (5.7), we have that

$$\begin{aligned} & P(\eta > [\kappa n]) \\ & = P(T_0 = 0)P(T_1 = 0 \mid T_0 = 0) \cdots P(T_{[\kappa n]} = 0 \mid T_0 = 0, \dots, T_{[\kappa n]-1} = 0) \\ & \leq P(T_0 = 0)(1 - \tilde{\gamma})^{[\kappa n]}. \end{aligned} \tag{5.9}$$

Now we are going to find an upper bound for  $E_x \rho^{\tau_{[\kappa n]} + n_0}$ .

Using the last-exit representation, we obtain

$$\begin{aligned} P\left(\tau_i - \tau'_{i-1} = k \mid \mathcal{B}_{i-1}\right) & = \sum_{s=0}^{\tau'_{i-1} - \tau_{i-1}} \int_K P_{X_{\tau_{i-1}}}^{\tau'_{i-1} - \tau_{i-1} - s}(dy) P_y(\tau_K = k + s) \\ & \leq \sup_{y \in K} \{P_y(\tau_K \geq k)\} \leq \rho^{-k} \sup_{y \in K} \{E_y \rho^{\tau_K}\}, \end{aligned}$$



which implies, for  $1 < r < \rho$ ,

$$\begin{aligned} E\left(r^{\tau_i - \tau'_{i-1}} \middle| \mathcal{B}_{i-1}\right) &= \sum_{k=0}^{\infty} r^k P(\tau_i - \tau'_{i-1} = k \mid \mathcal{B}_{i-1}) \\ &\leq \sup_{y \in K} \{E_y \rho^{\tau_K}\} \sum_{k=0}^{\infty} (r/\rho)^k =: \xi(r) < \infty. \end{aligned} \quad (5.10)$$

Analogously we obtain

$$E\left(r^{\tau'_i - \tau_i} \middle| \sigma(X_0, \dots, X_{\tau_i}, X'_0, \dots, X'_{\tau'_{i-1}})\right) \leq \xi(r), \quad (5.11)$$

which implies in conjunction with (5.10), for  $i \geq 2$ , that

$$\begin{aligned} &E\left(r^{\tau_i - \tau_{i-1}} \middle| \sigma(X_0, \dots, X_{\tau_{i-1}}, X'_0, \dots, X'_{\tau'_{i-2}})\right) \\ &= E\left(\left[r^{\tau'_{i-1} - \tau_{i-1}} E(r^{\tau_i - \tau'_{i-1}} \mid \mathcal{B}_{i-1})\right] \middle| \sigma(X_0, \dots, X_{\tau_{i-1}}, X'_0, \dots, X'_{\tau'_{i-2}})\right) \\ &\leq (\xi(r))^2. \end{aligned} \quad (5.12)$$

Next we bound  $E_x r^{\tau'_0}$ , where  $E_x$  refers to the initial condition  $X_0 = x$ . We can apply (5.11) if  $\{X'_t\}$  visited  $K$  before or at  $\tau_0$ . Let  $\tau'_{-1} = \inf_j \{X'_j \in K\}$ . [We have either  $\tau'_{-1} < \tau_0$  or  $\tau'_{-1} = \tau'_0$ .]

We have that

$$\begin{aligned} E_x r^{\tau'_0} &= E_x r^{\tau'_0} I(\tau'_0 = \tau'_{-1}) + E_x r^{\tau'_0} I(\tau'_{-1} < \tau_0) \\ &\leq E_\pi r^{\tau_K} + E_x \left[ r^{\tau'_0 - \tau_0} E\left(r^{\tau'_0 - \tau_0} \middle| \sigma(X_0, \dots, X_{\tau_0}, X'_0, \dots, X'_{\tau'_{-1}})\right) \right] \\ &\leq E_\pi r^{\tau_K} + \xi(r) E_x r^{\tau_K}. \end{aligned} \quad (5.13)$$

From (5.10), (5.12) and (5.13) we conclude

$$\begin{aligned} E_x r^{\tau_{[\kappa n]}} &= E_x \left[ r^{\tau_{[\kappa n]} - 1} E\left(r^{\tau_{[\kappa n]} - \tau_{[\kappa n]} - 1} \mid \mathcal{B}_{[\kappa n] - 1}\right) \right] \\ &\leq \dots \leq E_x \left[ r^{\tau'_0 - \tau_0} E(r^{\tau_1 - \tau'_0} \mid \mathcal{B}_0) \right] (\xi(r))^{2[\kappa n] - 2} \\ &\leq (E_\pi r^{\tau_K} + \xi(r) E_x r^{\tau_K}) (\xi(r))^{2[\kappa n] - 1}, \end{aligned} \quad (5.14)$$

which implies

$$\begin{aligned} P(\tau_{[\kappa n]} + n_0 > n) &\leq r^{n_0 - n} E_x r^{\tau_{[\kappa n]}} \\ &\leq r^{n_0 - n} (E_\pi r^{\tau_K} + \xi(r) E_x r^{\tau_K}) (\xi(r))^{2[\kappa n]} \\ &\leq r^{n_0} (E_\pi r^{\tau_K} + \xi(r) E_x r^{\tau_K}) \left( \xi(r)^{2\kappa} / r \right)^n. \end{aligned} \quad (5.15)$$

Choosing  $\kappa$  small enough, we obtain the assertion.  $\square$

*Proof of Theorem 3.2.* As already announced, we set  $X_0 \equiv X_0^* \equiv x_0$ , where  $x_0 \in K \cap \mathcal{X}_{T_0}$  (for  $T$  sufficiently large,  $K \cap T$  is nonempty). We pair the chains  $\{X_t\}$  and  $\{X_t^*\}$  in such a way that we have at each transition step a maximal coupling: Given

$X_{t-1} = x_{t-1}$  and  $X_{t-1}^* = x_{t-1}^*$ , then the joint distribution of  $X_t$  and  $X_t^*$  is chosen such that

$$\begin{aligned} P \left( X_t = X_t^* \mid X_{t-1} = x_{t-1}, X_{t-1}^* = x_{t-1}^* \right) \\ = \int p_{X_t^* | X_{t-1}^* = x_{t-1}^*}^*(z) \wedge p_{X_t | X_{t-1} = x_{t-1}}(z) dz. \end{aligned} \quad (5.16)$$

If  $x_{t-1} \in \mathcal{X}_T$ , then we obtain from (A4) that

$$\begin{aligned} & \int p_{X_t^* | X_{t-1}^* = x_{t-1}^*}^*(z) \wedge p_{X_t | X_{t-1} = x_{t-1}}(z) dz \\ &= 1 - \frac{1}{2} \int \left| p_{X_t^* | X_{t-1}^* = x_{t-1}^*}^*(z) - p_{X_t | X_{t-1} = x_{t-1}}(z) \right| dz \\ &= 1 - \tilde{O} \left( T^{-\vartheta}, T^{-\lambda} \right). \end{aligned}$$

This implies, by  $P(X_0, \dots, X_{n-2} \notin \mathcal{X}_T^c) = 1 - O(nT^{-\vartheta})$ , that

$$P \left( X_{n-1} \neq X_{n-1}^* \mid X_0 = X_0^* = x_0 \right) = \tilde{O} \left( nT^{-\vartheta}, T^{-\lambda} \right). \quad (5.17)$$

For the conditional densities of  $X_n$  and  $X_n^*$  we get

$$\begin{aligned} & \left| p_{X_n | X_0 = x_0}(z) - p_{X_n^* | X_0^* = x_0}^*(z) \right| \\ & \leq \int |p_\varepsilon(z - m(y)) - \hat{p}_\varepsilon(z - \widehat{m}(y))| \left[ p_{X_{n-1} | X_0 = x_0}(y) \wedge p_{X_{n-1}^* | X_0^* = x_0}^*(y) \right] dy \\ & \quad + \int p_\varepsilon(z - m(y)) \left( p_{X_{n-1} | X_0 = x_0}(y) - p_{X_{n-1}^* | X_0^* = x_0}^*(y) \right)_+ dy \\ & \quad + \int \hat{p}_\varepsilon(z - \widehat{m}(y)) \left( p_{X_{n-1}^* | X_0^* = x_0}^*(y) - p_{X_{n-1} | X_0 = x_0}(y) \right)_+ dy \\ & \leq \sup_z \{ \|p_\varepsilon(\cdot - m(z)) - \hat{p}_\varepsilon(\cdot - \widehat{m}(z))\|_\infty \} \\ & \quad + (\|p_\varepsilon\|_\infty + \|\hat{p}_\varepsilon\|_\infty) P \left( X_{n-1} \neq X_{n-1}^* \mid X_0 = X_0^* = x_0 \right) \\ & = \tilde{O} \left( nT^{-\vartheta}, T^{-\lambda} \right). \end{aligned} \quad (5.18)$$

Since  $\{X_t\}$  and  $\{X_t^*\}$  are geometrically ergodic, we obtain with  $n = [C_\lambda \log T]$  that

$$\begin{aligned} & |\pi(B) - \pi^*(B)| \\ & \leq |P^n(x_0, B) - \pi(B)| + |P^{*n}(x_0, B) - \pi^*(B)| \\ & \quad + |P^n(x_0, B) - P^{*n}(x_0, B)| \\ & = \tilde{O} \left( T^{-\lambda}, T^{-\lambda} \right) + \tilde{O} \left( \lambda(B) T^{-\vartheta}, T^{-\lambda} \right) \end{aligned} \quad (5.19)$$

holds simultaneously over all measurable  $B$ .  $\square$

Before we turn to the next proofs, we quote a useful lemma from Neumann and Kreiss (1997). This lemma describes the stochastic behavior of sums of geometrically  $\beta$ -mixing random variables.

**Lemma 5.1.** *Suppose that  $(Z_t)_{t=1, \dots, T}$  is geometrically  $\beta$ -mixing and  $EZ_t = 0$ .*

(i) If  $|Z_t| \leq 1$  almost surely, then

$$\sum_{t=1}^T Z_t = \tilde{O} \left( \min \left\{ \sqrt{T \log T}, \sqrt{\sum_t \text{var}(Z_t) \log T} + (\log T)^2 \right\}, T^{-\lambda} \right).$$

(ii) Under the weaker assumption that  $\forall M < \infty \exists C_M < \infty$  such that  $E|Z_t|^M \leq C_M$  we have

$$\sum_{t=1}^T Z_t = \tilde{O} \left( \sqrt{\sum_t \text{var}(Z_t) \log T} + T^\delta, T^{-\lambda} \right).$$

*Proof of Lemma 3.1.* First, we choose two constants  $0 < \delta' < \delta''$  with  $N/T = o(T^{-2\delta''})$  and define

$$\mathcal{X}_T = \left\{ x \in [-T^{\delta'}, T^{\delta'}] \mid \pi(x) \geq T^{-\delta''} \right\}. \quad (5.20)$$

(i) *Proof of  $\pi(\mathcal{X}_T^c) = O(T^{-\vartheta})$*

It is obvious that

$$\pi(\mathcal{X}_T^c \cap [-T^{\delta'}, T^{\delta'}]) = O(T^{-\vartheta}). \quad (5.21)$$

Hence, it remains to show that

$$\pi([-T^{\delta'}, T^{\delta'}]^c) = O(T^{-\lambda}). \quad (5.22)$$

According to (A3)(i), we have that

$$|m(x)| \leq \alpha|x| + C_1$$

holds for all  $x$  and some  $\alpha < 1$ . Hence, we get the estimate

$$\begin{aligned} |X_t| &\leq \alpha|X_{t-1}| + (C_1 + |\varepsilon_t|) \\ &\leq \dots \leq \alpha^t|X_0| + \alpha^{t-1}(C_1 + |\varepsilon_1|) + \dots + (C_1 + |\varepsilon_t|) \\ &\stackrel{d}{=} \alpha^t|X_0| + \sum_{s=1}^t \alpha^{s-1}(C_1 + |\varepsilon_s|). \end{aligned}$$

Setting for a moment  $X_0 \equiv 0$ , we obtain by (A6) that

$$\begin{aligned} \int |x|^M \pi(x) dx &= \int |x|^M \lim_{t \rightarrow \infty} p_{X_t}(x) dx \\ &\leq \int \sup_t \{p_{X_t}(x) |x|^M\} dx \\ &\leq E \left( \sum_{t=1}^{\infty} \alpha^{t-1} (C_1 + |\varepsilon_t|) \right)^M \\ &\leq E \left( \sum_{t=1}^{\infty} \alpha^{t-1} (C_1 + |\varepsilon_1|) \right)^M < \infty, \end{aligned}$$

which implies (5.22) by Markov's inequality.

(ii) *Estimation of  $\sup_{x \in \mathcal{X}_T} \{|\widehat{m}_N(x) - m(x)|\}$*

Define the intervals

$$I_{kl} = [F_X^{-1}(k/T), F_X^{-1}(l/T)], \quad 0 \leq k < l \leq T,$$

where  $F_X(t) = \int_{-\infty}^t \pi(x) dx$  is the cumulative distribution function of  $X_t$ . We obtain from Lemma 5.1 that

$$\begin{aligned} & \#\{1 \leq t \leq T : X_{t-1} \in I_{kl}\} = (l - k) \\ & = \tilde{O}\left(\min\left\{\sqrt{l-k} \log T + (\log T)^2, \sqrt{T \log T}\right\}, T^{-\lambda}\right). \end{aligned} \quad (5.23)$$

Let

$$\mathcal{N}_{N,T}(x) = [x - n_{N,T}(x), x + n_{N,T}(x)]$$

be the nonrandom counterpart to  $\widehat{\mathcal{N}}_N(x)$ , where  $n_{N,T}(x)$  is the minimal number such that  $\pi(\mathcal{N}_{N,T}(x)) = N/T$ . We obtain from (5.23) that

$$P\left(\mathcal{N}_{N-C_\lambda\sqrt{N}\log T,T}(x) \subseteq \widehat{\mathcal{N}}_N(x) \subseteq \mathcal{N}_{N+C_\lambda\sqrt{N}\log T,T}(x) \quad \text{for all } x \in \mathbb{R}\right) \geq 1 - O(T^{-\lambda}), \quad (5.24)$$

which implies, by  $\mathcal{N}_{N+C_\lambda\sqrt{N}\log T,T}(x) \subseteq [x - CT^{\delta''}N/T, x + CT^{\delta''}N/T]$ ,

$$\frac{1}{N} \sum_{X_{t-1} \in \widehat{\mathcal{N}}_N(x)} m(X_{t-1}) - m(x) = \tilde{O}\left(T^{\delta''}N/T, T^{-\lambda}\right). \quad (5.25)$$

Further, we have by Lemma 5.1

$$\begin{aligned} & \sup_x \left\{ \left| \sum_{t: X_{t-1} \in \widehat{\mathcal{N}}_N(x)} \varepsilon_t \right| \right\} \\ & \leq \sup_x \left\{ \left| \sum_{t: X_{t-1} \in \mathcal{N}_{N,T}(x)} \varepsilon_t \right| \right\} + \sup_x \left\{ \sum_{t: X_{t-1} \in \widehat{\mathcal{N}}_N(x) \Delta \mathcal{N}_{N,T}(x)} |\varepsilon_t| \right\} \\ & = \tilde{O}\left(\sqrt{N} \log T, T^{-\lambda}\right), \end{aligned} \quad (5.26)$$

which yields the assertion.  $\square$

*Proof of Lemma 3.2.* We define

$$\tilde{p}_\varepsilon(x) = \frac{1}{Th} \sum_{t=1}^T K\left(\frac{x - \varepsilon_t}{h}\right)$$

and

$$\overline{m}(x) = \left( \int_{\mathcal{N}_{N,T}(x)} \pi(x) dx \right)^{-1} \int_{\mathcal{N}_{N,T}(x)} m(x) \pi(x) dx.$$

The difference of  $\hat{p}_\varepsilon(x)$  and  $p_\varepsilon(x)$  will be decomposed as follows:

$$\begin{aligned}
& \hat{p}_\varepsilon(x) - p_\varepsilon(x) \\
&= \frac{1}{Th} \sum_{t: X_{t-1} \in [-T^{\delta'}, T^{\delta'}]} \left[ K\left(\frac{x - \hat{\varepsilon}_t}{h}\right) - K\left(\frac{x - [X_t - \overline{m}(X_{t-1})]}{h}\right) \right] \\
&\quad + \frac{1}{Th} \sum_t \left\{ \left[ K\left(\frac{x - [X_t - \overline{m}(X_{t-1})]}{h}\right) - K\left(\frac{x - \varepsilon_t}{h}\right) \right] I(X_{t-1} \in [-T^{\delta'}, T^{\delta'}]) \right. \\
&\quad \quad \left. - E \left[ K\left(\frac{x - [X_t - \overline{m}(X_{t-1})]}{h}\right) - K\left(\frac{x - \varepsilon_t}{h}\right) \right] I(X_{t-1} \in [-T^{\delta'}, T^{\delta'}]) \right\} \\
&\quad + \frac{1}{Th} \sum_t E \left[ K\left(\frac{x - [X_t - \overline{m}(X_{t-1})]}{h}\right) - K\left(\frac{x - \varepsilon_t}{h}\right) \right] I(X_{t-1} \in \mathcal{X}_T) \\
&\quad + \frac{1}{Th} \sum_t E \left[ K\left(\frac{x - [X_t - \overline{m}(X_{t-1})]}{h}\right) - K\left(\frac{x - \varepsilon_t}{h}\right) \right] I(X_{t-1} \in [-T^{\delta'}, T^{\delta'}] \setminus \mathcal{X}_T) \\
&\quad + \frac{1}{Th} \sum_{t: X_{t-1} \notin [-T^{\delta'}, T^{\delta'}]} \left[ K\left(\frac{x - \hat{\varepsilon}_t}{h}\right) - K\left(\frac{x - \varepsilon_t}{h}\right) \right] \\
&\quad + \tilde{p}_\varepsilon(x) - p_\varepsilon(x) \\
&= T_1(x) + \dots + T_6(x).
\end{aligned} \tag{5.27}$$

It can be shown (see also (5.26)) that

$$\sup_{y \in [-T^{\delta'}, T^{\delta'}]} \{ |\widehat{m}_N(y) - \overline{m}(y)| \} = \tilde{O}\left(N^{-1/2} T^\delta, T^{-\lambda}\right),$$

which implies

$$\sup_x \{|T_1(x)|\} = \tilde{O}\left(h^{-1} N^{-1/2} T^\delta, T^{-\lambda}\right). \tag{5.28}$$

Further, we can show for *fixed*  $x$  that

$$|T_2(x)| + |T_6(x)| = \tilde{O}\left((Th)^{-1/2} \log T, T^{-\lambda}\right) + O(h^2)$$

holds. By proving this result on a sufficiently fine grid and using that  $P(|\varepsilon_t| > T^\delta) = O(T^{-\lambda})$  we get

$$\sup_x \{|T_2(x)| + |T_6(x)|\} = \tilde{O}\left((Th)^{-1/2} \log T, T^{-\lambda}\right) + O(h^2). \tag{5.29}$$

Since  $p_\varepsilon$  is Lipschitz, we have

$$\sup_x \{|T_3(x)|\} = O\left(\sup_{y \in \mathcal{X}_T} \{|\overline{m}(y) - m(y)|\}\right) = O\left(T^\delta N/T\right). \tag{5.30}$$

By (5.23) we obtain

$$\sup_x \{|T_4(x)|\} = O\left(P(X_{t-1} \in [-T^{\delta'}, T^{\delta'}] \setminus \mathcal{X}_T)\right) = O\left(T^{-\vartheta}\right) \tag{5.31}$$

and, by (5.24),

$$\sup_x \{|T_5(x)|\} = O(T^{-\vartheta}). \quad (5.32)$$

From (5.27) to (5.32) we obtain (i).

To get (ii), we first conclude from (i) that

$$\int_{[-T^\delta, T^\delta]} |\hat{p}_\varepsilon(x) - p_\varepsilon(x)| dx = O(T^{-\vartheta}).$$

Since, furthermore,  $\int_{[-T^\delta, T^\delta]^c} p_\varepsilon(x) dx = O(T^{-\lambda})$ , we obtain (ii).  $\square$

*Proof of Lemma 4.1.* In order to save space, we give only a brief sketch of the proof. We assume throughout this proof that  $(X_0, \dots, X_T) \in \Omega_T$ .

We consider “small” intervals  $I_k = [(k-1)g, kg)$ , where an appropriate choice of  $g$  will become clear from the calculations below. To construct a pairing of  $X_0, \dots, X_T$  and  $X_0^*, \dots, X_T^*$  such that  $\sup_x \{\sum_t K((x - X_{t-1})/h)\varepsilon_t - \sum_t K((x - X_{t-1}^*)/h)\varepsilon_t^*\}$  is small, we try first to find a pairing such that the partial sums w.r.t. the  $I_k$ ,

$$Z_k = \sum_{t: X_{t-1} \in I_k} \varepsilon_t$$

and

$$Z_k^* = \sum_{t: X_{t-1}^* \in I_k} \varepsilon_t^*,$$

are close to each other. This will be achieved by a simultaneous embedding of  $\varepsilon_1, \dots, \varepsilon_T$  and  $\varepsilon_1^*, \dots, \varepsilon_T^*$  in a common set of independent Wiener processes  $W_k$  assigned to these intervals.

(i) *Embedding of  $\varepsilon_1, \dots, \varepsilon_T$  and  $\varepsilon_1^*, \dots, \varepsilon_T^*$*

In order to embed the  $\varepsilon_t$  in the Wiener processes, we define appropriate stopping times  $\tau_k^{(t)}$ . To initialize this procedure, we set  $\tau_k^{(0)} = 0$  for all  $k$ . Let  $k_1$  be the index of that interval  $I_k$  into which  $X_0$  did fall. Then we embed  $\varepsilon_1$  in the Wiener process  $W_{k_1}$ , that is, according to Lemma A.2 in Hall and Heyde (1980, Appendix A.1) we choose an appropriate stopping time  $\tau_{k_1}^{(1)}$  such that

$$\varepsilon_1 = W_{k_1}(\tau_{k_1}^{(1)}) - W_{k_1}(\tau_{k_1}^{(0)}).$$

For  $k \neq k_1$ , we set  $\tau_k^{(1)} = \tau_k^{(0)}$ .

Now we repeat this procedure to embed successively  $\varepsilon_2, \varepsilon_3, \dots, \varepsilon_T$ . Assume that  $\varepsilon_1, \dots, \varepsilon_{t-1}$  are already embedded. Let  $k_t$  be the index of that interval into which  $X_{t-1}$  did fall. Then we use the remaining part of the process  $W_{k_t}$  to represent  $\varepsilon_t$ , that is we choose a stopping time  $\tau_{k_t}^{(t)}$  such that

$$\varepsilon_t = W_{k_t}(\tau_{k_t}^{(t)}) - W_{k_t}(\tau_{k_t}^{(t-1)}).$$

For  $k \neq k_t$ , we set  $\tau_k^{(t)} = \tau_k^{(t-1)}$ .

After  $T$  such steps we obtain the representation

$$Z_k = W_k(\tau_k^{(T)}).$$

Now we use the same Wiener processes to embed successively  $\varepsilon_1^*, \dots, \varepsilon_T^*$ . By exactly the same steps as above we obtain

$$Z_k^* = W_k(\tilde{\tau}_k^{(T)}).$$

(ii) *Closeness of  $Z_k$  and  $Z_k^*$*

The proof of the closeness of  $Z_k$  and  $Z_k^*$  will be based on an upper estimate of  $|\tau_k^{(T)} - \tilde{\tau}_k^{(T)}|$ . First of all, note that  $\tau_k^{(t)} - \tau_k^{(t-1)}$  depends only on  $X_{t-1}$  and  $\{W_{k_t}(s), \tau_{k_t}^{(t-1)} \leq s \leq \tau_{k_t}^{(t)}\}$ . Hence, these differences are also geometrically  $\beta$ -mixing. Therefore, we obtain from Lemma 5.1(ii) that

$$|\tau_k^{(T)} - E\tau_k^{(T)}| + |\tilde{\tau}_k^{(T)} - E\tilde{\tau}_k^{(T)}| = \tilde{O}\left((Tg)^{1/2} \log(T) + T^\delta, T^{-\lambda}\right).$$

Moreover, we get from Theorem 3.2 that

$$E\tau_k^{(T)} - E\tilde{\tau}_k^{(T)} = T \left( EI(X_0 \in I_k) \varepsilon_1^2 - EI(X_0^* \in I_k) \varepsilon_1^{*2} \right) = O\left(T^{1-\kappa} g\right).$$

Provided that  $1/g = O(T^{1-\kappa})$ , these two estimates yield that

$$P\left(|\tau_k^{(T)} - \tilde{\tau}_k^{(T)}| > TgT^{-\kappa}\right) = O\left(T^{-\lambda}\right)$$

holds for some  $\kappa > 0$ . This implies, by Lemma 1.2.1 in Csörgő and Révész (1981, p. 29), that

$$P\left(|Z_k - Z_k^*| > (Tg)^{1/2} T^{-\kappa}\right) = O\left(T^{-\lambda}\right).$$

With a simple extra argument we can also prove that

$$P\left(|Z_k - Z_k^*| > (Tg)^{1/2} T^{-\kappa} \text{ for all } k\right) = O\left(T^{-\lambda}\right). \quad (5.33)$$

(iii) *Closeness of  $\sum_t K((x - X_{t-1})/h) \varepsilon_t$  and  $\sum_t K((x - X_{t-1}^*)/h) \varepsilon_t^*$*

To exploit the result of our partial sum approximation, we approximate  $K((x - \cdot)/h)$  by piecewise constant functions, i.e.,

$$K((x - y)/h) = \sum_k \alpha_k(x) I(y \in I_k) + r(x, y),$$

where

$$\begin{aligned} \alpha_k(x) &= \frac{1}{g} \int_{I_k} K((x - y)/h) dy, \\ r(x, y) &= K((x - y)/h) - \sum_k \alpha_k(x) I(y \in I_k). \end{aligned}$$

This yields the decomposition

$$\begin{aligned}
& \left| \sum_t K((x - X_{t-1})/h) \varepsilon_t - \sum_t K((x - X_{t-1}^*)/h) \varepsilon_t^* \right| \\
& \leq \sum_k |\alpha_k(x)| |Z_k - \tilde{Z}_k| + \left| \sum_t r(x, X_{t-1}) \varepsilon_t \right| + \left| \sum_t r(x, X_{t-1}^*) \varepsilon_t^* \right| \\
& = U_1(x) + U_2(x) + U_3(x).
\end{aligned} \tag{5.34}$$

Now we obtain by (5.33) that

$$\sup_x \{|U_1(x)|\} = \tilde{O}\left((Th)^{1/2} T^{-\kappa} \sqrt{h/g}, T^{-\lambda}\right). \tag{5.35}$$

Since  $K$  is Lipschitz we have  $\sup_{x,y} \{|r(x,y)|\} = O(g/h)$ , which implies by Lemma 5.1 that

$$|U_2(x)| + |U_3(x)| = \tilde{O}\left((Th)^{1/2} g/h \log(T) + T^\delta, T^{-\lambda}\right). \tag{5.36}$$

Let

$$\mathcal{X}_0 = \left\{ x \mid \int_{\text{supp}(K((x-\cdot)/h))} \pi(x) + \pi^*(x) dx \geq T^{-1} \right\}.$$

It is clear that there exists some set  $\mathcal{K}_0$  of cardinality  $O(T^C)$  such that  $\mathcal{X}_0 \subseteq \bigcup_{k \in \mathcal{K}_0} I_k$ . Proving (5.36) for  $x$  on a sufficiently fine grid on  $\bigcup_{k \in \mathcal{K}_0} I_k$  we can show that

$$\sup_{x \in \mathcal{X}_0} \{|U_2(x)| + |U_3(x)|\} = \tilde{O}\left((Th^{1/2} g/h \log(T) + T^\delta, T^{-\lambda}\right). \tag{5.37}$$

Finally, one can show by simple extra arguments that

$$\sup_{x \in \mathbb{R} \setminus \mathcal{X}_0} \{|U_2(x)| + |U_3(x)|\} = \tilde{O}\left(T^\delta, T^{-\lambda}\right). \tag{5.38}$$

The assertion follows now from (5.34), (5.35), (5.37) and (5.38).  $\square$

*Proof of Lemma 4.2.* We will show that

$$\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_t \left( X_t^* - \vartheta^* m_o(X_{t-1}^*) \right) m_o(X_{t-1}^*) \\
& = \frac{1}{\sqrt{T}} \sum_t \left( X_t^* m_o(X_{t-1}^*) - E^*[X_1^* m_o(X_0^*)] \right) \\
& \quad - \vartheta^* \frac{1}{\sqrt{T}} \sum_t \left( m_o(X_{t-1}^*)^2 - E^*[m_o(X_0^*)^2] \right) \\
& \Rightarrow \mathcal{N}(0, \tau^2).
\end{aligned} \tag{5.39}$$

and

$$\frac{1}{T} \sum_t m_o(X_{t-1}^*)^2 \rightarrow E m_o(X_0)^2,$$

if  $(X_0, \dots, X_T) \in \Omega_T$ . Since the method of proof is simple we demonstrate the argument by showing weak convergence of (5.39), only. Concerning this we make use of a CLT for a triangular array of strongly mixing random variables given in Politis,



Romano and Wolf (1997, Theorem A1). Since the bootstrap process is assumed to be stationary the assumptions of Politis, Romano and Wolf (1997) simplify to

- (i)  $E^*(X_1^* m_o(X_0^*))^{2+2\delta} \leq \Delta$  for all  $T$  and some  $\delta > 0$ ,
- (ii) there exists  $\sigma^2 \in (0, \infty)$  such that for  $T \rightarrow \infty$

$$\text{Var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T X_t^* m_o(X_{t-1}^*) \right) \rightarrow \sigma^2 ,$$

- (iii)  $\sum_{k=0}^{\infty} (k+1)^2 \alpha_T^{\delta/(4+\delta)}(k) \leq K$  for all  $T$ .

Here  $\alpha_T(\cdot)$  denote the strong mixing coefficients of the bootstrap process.

Since all moments of the bootstrap process exist and are uniformly bounded (use (A4)(v) and the same argument as in the proof of Lemma 3.1) we obtain (i) because of (A7).

(iii) is an immediate consequence of Theorem 3.1 and the inequality  $|\alpha_T(\cdot)| \leq |\beta^*(\cdot)|$ . To see (ii), let

$$\sigma^2 = \text{Var} (X_1 m_o(X_0)) + 2 \cdot \sum_{k=1}^{\infty} \text{Cov} (X_1 m_o(X_0), X_{k+1} m_o(X_k)) .$$

Thus, it suffices to prove (as  $T \rightarrow \infty$ )

$$(a) \ E^* X_1^* m_o(X_0^*) \rightarrow E X_1 m_o(X_0)$$

and

$$(b) \ E^* X_{k+1}^* m_o(X_k^*) X_1^* m_o(X_0^*) \rightarrow E X_{k+1} m_o(X_k) X_1 m_o(X_0), \quad k = 0, 1, 2, \dots$$

The left-hand side of (b) equals

$$\begin{aligned} & \int_{\mathbb{R}^{k+2}} x_{k+1} m_o(x_k) x_1 m_o(x_0) \hat{p}_\varepsilon(x_{k+1} - \hat{m}(x_k)) \dots \hat{p}_\varepsilon(x_1 - \hat{m}(x_0)) \pi^*(x_0) dx_{k+1} \dots dx_0 \\ &= \int_{X_T^{k+2}} x_{k+1} m_o(x_k) x_1 m_o(x_0) \hat{p}_\varepsilon(x_{k+1} - \hat{m}(x_k)) \dots \hat{p}_\varepsilon(x_1 - \hat{m}(x_0)) \pi^*(x_0) dx_{k+1} \dots dx_0 + o(1) \\ &= E X_{k+1} m_o(X_k) X_1 m_o(X_0) + o(1) , \end{aligned}$$

because of (A4)(ii), (iii) and Theorem 3.2.

The argument for (a) is quite similar and therefore omitted. □

## REFERENCES

- Athreya and Ney. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245**, 493–501.
- Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Lecture Notes in Statistics* **110**, Springer, New York.
- Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.
- Davydov, Yu. A. (1973). Mixing conditions for Markov chains. *Theory Probab. Appl.* **18**, 312–328.
- Doukhan, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**, Springer, New York.
- Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré* **31**, 393–427.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1–26.

- Efron, B. and Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Science* **1**, 54–77.
- Franke, J., Kreiss, J.-P. and Mammen, E. (1997). Bootstrap of kernel smoothing in nonlinear time series. Discussion Paper 20/97, SFB 373, Humboldt University, Berlin.
- Franke, J. and Wendel, M. (1992). A bootstrap approach for nonlinear autoregressions - some preliminary results. In: Jöckel, K. H., Rothe, G. and Sendler, W. eds.: *Bootstrapping and Related Techniques*. Lecture Notes in Economics and Mathematical Systems 376, Springer, Berlin Heidelberg.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.
- Holbert, D. and Son, M. S. (1986). Bootstrapping a time series model: Some empirical results. *Comm. Statist. A* **15**, 3669–3691.
- Kreiss, J.-P., Neumann, M. H. and Yao, Q. (1998). Bootstrap tests for simple structures in nonparametric time series regression, Preprint No. 98/07, TU Braunschweig.
- Kreutzberger, E. (1993). Bootstrap für nichtlineare AR(1)-Prozesse. Thesis, University of Kaiserslautern, Germany.
- Lindvall, T. (1992). *Lectures on the Coupling Method*. Springer, New York.
- Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- Neumann, M. H. (1996). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.*, to appear.
- Neumann, M. H. (1997). On robustness of model-based bootstrap schemes in nonparametric time series analysis, Discussion Paper 88/97, SFB 373, Humboldt University, Berlin.
- Neumann, M. H. and Kreiss, J.-P. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.*, to appear.
- Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **43**, 309–318.
- Nummelin, E. and Tuominen, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proc. Appl.* **12**, 187–202.
- Politis, D. N., Romano, J. P. and Wolf, M. (1997). Subsampling for heteroskedastic time series. *J. Econometrics* **81**, 281–317.
- Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4**, 185–207.
- Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Jürgen Franke  
 Fachbereich Mathematik  
 Universität Kaiserslautern  
 Erwin-Schrödinger-Straße  
 D – 67663 Kaiserslautern  
 Germany  
 E-mail: franke@mathematik.uni-kl.de

Jens-Peter Kreiss  
 Fachbereich Mathematik  
 Technische Universität Braunschweig  
 Pockelsstraße 14  
 D – 38106 Braunschweig  
 Germany  
 E-mail: j.kreiss@tu-bs.de

Enno Mammen  
 Institut für Angewandte Mathematik  
 Universität Heidelberg  
 Im Neuenheimer Feld 294  
 D – 69120 Heidelberg  
 Germany  
 E-mail: enno@statlab.uni-heidelberg.de

Michael H. Neumann  
 Sonderforschungsbereich 373  
 Humboldt-Universität zu Berlin  
 Spandauer Straße 1  
 D – 10178 Berlin  
 Germany  
 E-mail: neumann@wiwi.hu-berlin.de